

足球博彩市场建模及市场有效性假说检验

黄晓雯

(厦门大学 经济学院 福建 厦门 361005)

【摘要】在金融市场上由于资产的内在价值无法确切估量,因此对市场有效性假说的验证存在很大分歧。而在博彩市场,博彩商结合所有能获得的信息给出不同体育比赛的赔率(当前市场价格),不同信念的“投资者”通过对最终比赛结果(内在价值)的预测来进行交易,为市场有效性假说提供了十分恰当的检验平台。文章利用近十年来英超联赛数据,通过建立比赛胜负结果和比分的计量模型与检验,发现足球博彩市场不存在系统性的定价偏差,总体上符合弱势有效假说,在存在交易费用的基础上可能正收益的交易策略。

【关键词】足球博彩;赔率;市场有效性;泊松回归;logistic 选择模型

【中图分类号】F224.0;G80-05 【文献标识码】A 【文章编号】1004-2768(2012)08-0119-03

一、问题的提出及文献综述

作为现代金融学的两大基石之一,有效市场假说自提出以来一直在证券市场理论中占据主导地位。长期以来,国内研究一直致力于通过金融市场检验市场有效性假说。但现实中由于金融市场中资产真实价格不可知性的存在,直接从价格来检验市场有效性几乎是不可能的。博彩市场却为从价格角度直接检验市场有效性假说提供了一个有效的平台。相比于一直活跃在金融市场中的股票价格,博彩市场中每一场赌注都有一项明确的最终价格,另外博彩商事前根据所有可获得信息开出的赔率可视为资产的事前价格,也就是说在博彩市场中资产真实价格和市场价格都是可知的,这就为从价格角度直接测算市场有效性提供了可能,由此,Thaler 和 Ziemba(1998)指出博彩市场比股票市场更适合用于检验市场有效性和理性预期。

本文重点考虑固定赔率足球博彩,在固定赔率足球博彩中,博彩公司根据所能获得的所有公开市场有效信息开出胜平负相应赔率,彩民根据博彩公司开出的赔率对比赛结果的形成的主观判断,对主客队的胜平负下注。所谓赔率即博彩公司对胜平负结果开出一组赔付系数,赢面大的一方,相应的赔率自然就低,赢面小的一方,其赔率就相对较高。

关于足球博彩业,国外文献做了大量的相关研究。Jaffe 和 Winkler(1976)指出足球博彩市场与证券市场之间存在众多的相似性,Ruhm(2003)解释了一笔简单的投注如何类比于金融市场中的一项期权。关于市场有效性检验,Thaler 和 Ziemba(1988)指出在强势有效市场中所有的投注都应该有同等期望收益。有效性检验的最基本途径为通过博彩商给出的赔率对比赛结果进行预测,若不存在系统性偏差则证明市场有效,在此基础上各种研究通过在简单回归中加入其他变量扩展和深入模型,Gray 和 Gray(1997)在回归模型中加入比赛记录变量如最近四场比赛中主队赢的场数等,发现部分记录变量对结果预测影响显著,得出了市场存在一定程度非有效的结论。Golec 和 Tamarkin(1991)对回归模型也做了类似修正。

对比于国外的大量研究,国内却几乎没有这方面的研究,鉴于博彩市场更适用于有效性假说检验的种种特征,本文从足球博彩市场入手,对市场有效性假说进行了检验,利用 logistic 离散选择模型,对 2012 赛季的比赛结果进行了预测,进一步地对可获利性策略进行了探索,通过泊松模型估计了各队内在参数,并对比赛进球数结果进行了预测。

二、博彩市场有效性检验

为了检验最终价格中是否包含了市场中所有公开的有效信息,可通过事前信息的集合——赔率对比赛结果的预测的有效性来衡量。由此建立系列回归模型:

$$P_{home}=a_0+a_1y_{home}+\varepsilon_1$$

$$P_{draw}=b_0+b_1y_{draw}+\varepsilon_2$$

$$P_{away}=c_0+c_1y_{away}+\varepsilon_3$$

其中 P_{home} 、 P_{draw} 、 P_{away} 分别表示最终比赛结果中主队胜、平、负的概率, y_{home} 、 y_{draw} 、 y_{away} 分别表示赔率中隐含的事前比赛胜平负结果,赔率越高意味着该队获胜的预期概率越小,由于博彩公司开出的赔率中包含了利润,因此用赔率转化成的胜平负概率之和大于 1,因此此处需进行归一化调整,调整后得到 y 。通过系列联合检验 $H_0^{(1)}:a_0=0$ 且 $a_1=1$, $H_0^{(2)}:b_0=0$ 且 $b_1=1$, $H_0^{(3)}:c_0=0$ 且 $c_1=1$,对市场有效性进行判断,若原假设成立,则赔率对比赛结果预测力强,即赔率中包含了大量的可获得信息,也就意味着市场有效性假说得到验证。

表 1 2002—2012 英格兰足球超级联赛数据回归结果

$P_{home}=a_0+a_1y_{home}+\varepsilon_1$			
a_0	a_1	联合检验 F 值	R^2
0.002901	1.029077	0.066419	0.787804
$P_{draw}=b_0+b_1y_{draw}+\varepsilon_2$			
b_0	b_1	联合检验 F 值	R^2
0.00084	0.904927	0.033529	0.208703
$P_{away}=c_0+c_1y_{away}+\varepsilon_3$			
c_0	c_1	联合检验 F 值	R^2
-0.044922	1.073207	0.3924057	0.889071

注:假设 t 统计量不相关, $F=(t_1^2+t_2^2)/2$

【收稿日期】2012-06-22

【作者简介】黄晓雯,女,福建泉州人,厦门大学经济学院,研究方向:统计学。

回归结果显示, 检验的 F 值都较小, 接受原假设 $H_0^{(1)}$ 、 $H_0^{(2)}$ 、 $H_0^{(3)}$, 市场在该检验下有效, 也就是说赔率有效地囊括了过去所有可获得的信息, 并将这些信息综合起来对最终比赛结果做出了准确的预测, 根据 Kuypers (2000) 的结论, 更准确地说这种结果为弱势市场有效性提供了有力的证据。根据有效市场假设, 此时对所有理性的参与者来说, 市场中不存在能够获取任何超额利润的交易策略, 因为所有的信息已反映在当前的赔率中。

三、比赛结果预测及投注策略探索

(一) 三元离散选择模型建立

为了进一步有效地利用赔率对具体某一场的比赛结果做出预测, 建立三元 logistic 选择模型, 自变量为由胜赔、平赔、负赔计算所得事前概率, 分别记 x_1 、 x_2 、 x_3 , 因变量为比赛结果 D_i , 为离散选择变量, 根据主队胜平负分别取之 1, 0, -1。为了探索模型的最优形式, 以逐步进入的方式对模型进行筛选, 结果表明, 由于 x_3 的进入, x_1 和 x_2 的贡献由显著变为不显著, 可能因为三个变量之间多重共线性的存在, 因此剔除胜赔概率和平赔概率, 模型的最后估计结果如下。与预期一致的, 负赔概率 x_3 每变动一个单位, 客队胜的概率相对于主队胜的概率平均变动 $e^{6.773}$ 倍, 负赔概率与客队胜的概率同方向变化。

$$\ln \frac{P_{-1}}{P_1} = -2.851 + 6.906x_3$$

$$\ln \frac{P_0}{P_1} = -1.605 + 3.451x_3$$

(二) 模型预测准确度和交易策略探索

使用 11 年历史数据所建模型对 2012 年 380 场比赛的结果进行预测, 在多项 logistic 选择模型中, 由模型所得多个 logit 方程系数, 可对各类因变量概率进行估计。仅就主队胜的概率而言对预测结果进行分类, 并计算其占比, 将预测结果与实际比赛结果进行对比。

表 2 logistic 预测概率结果和预测准确度

Logistic 预测概率	频数	所占比例 (%)	预测准确度 (%)
$P_1 < 0.425$	126	0.33	0.46
$0.425 < P_1 < 0.475$	37	0.10	0.38
$0.475 < P_1 < 0.525$	47	0.12	0.38
$0.525 < P_1 < 0.575$	33	0.09	0.52
$P_1 > 0.575$	137	0.36	0.66

注: 利用以上公式, 表中模型所得预测概率计算公式分别为:

$$P_{-1} = \frac{\exp(-2.851 + 6.906x_3)}{\exp(-2.851 + 6.906x_3) + \exp(-1.605 + 3.451x_3) + 1}$$

$$P_0 = \frac{\exp(-1.605 + 3.451x_3)}{\exp(-2.851 + 6.906x_3) + \exp(-1.605 + 3.451x_3) + 1}$$

$$P_1 = \frac{1}{\exp(-2.851 + 6.906x_3) + \exp(-1.605 + 3.451x_3) + 1}$$

显然, 对于主队胜的概率高的准确度相对也较高, 可达到 66%, 也就是说当比赛实力较为悬殊的时候, 模型能够对比赛的结果做出相当高的预测, 就 2012 年总体的预测效果来说, 预测的准确度达到了 54.55%, 预测效果较为理想。

另外, 对于彩民来说, 更关心的应该是所用模型能否用于制定可获利的交易策略, 根据模型给出的结果, 若预测主队胜则下注主队, 若预测客队胜则下注客队, 若预测与比赛结果一致, 则赚取相应赔率, 若预测与比赛结果不一致, 则失去该单位赌注, 由此计算上述区间内的平均收益率, 结果如表 3 所示。尽管在各个预测概率区间中模型的期望收益均为负, 但是在高预测概率段, 该交易策略的期望收益向 0 回归, 也就是说, 在这种情况下若将交易费用考虑在内, 该模型可能给出正收益策略。

四、进球数建模与结果预测

(一) 模型建立及独立性检验

在博彩市场, 除了固定赔率的方式, 彩民还可以直接就比

赛结果的进球数进行押注。

表 3 交易策略的平均收益

Logistic 预测概率区间	期望收益
$P_1 < 0.425$	-11.75%
$0.425 < P_1 < 0.475$	-11.16%
$0.475 < P_1 < 0.525$	-18.62%
$0.525 < P_1 < 0.575$	-0.18%
$P_1 > 0.575$	-5.26%

关于进球数的预测, Maher (1982) 在主客队进球数分布假设前提下提出了预测模型。该模型假设主客队进球均为独立的泊松分布, 参数值由双方队伍各自的进攻力和防守力决定, 对于主队来说还与主场效应有关。具体而言, 在 Maher 的模型中, 假设比赛双方为队伍 i 和队伍 j , 记 X_{ij} 和 Y_{ij} 分别为主队和客队比赛最终结果的进球数, α_1 和 α_2 分别为主队和客队的进攻力, β_1 和 β_2 分别为主队和客队的防守力, γ 表示主场效应, 其中 $\alpha > 0, \beta > 0, \gamma > 0$ 。则 X_{ij} 和 Y_{ij} 服从以下分布:

$$X_{ij} \sim \text{poisson}(\lambda_{ij})$$

$$Y_{ij} \sim \text{poisson}(\lambda_{ij})$$

首先需要对独立性假设做出检验。对于每一对比分 $i:j$ 计算 $\frac{f(i,j)}{f(i)*f(j)}$ 检验相关性, 检验结果如表 4 所示。结果显示, 主客队比分相对独立, 除了个别异常值, 比分基本上呈现出较强的独立性。

表 4 独立性检验结果

主队 \ 客队	0	1	2	3	4	5
0	1.01	0.90	1.03	1.05	1.69	2.06
1	0.96	1.04	1.00	1.04	0.96	0.80
2	1.00	1.05	1.04	0.85	0.51	0.44
3	1.03	1.05	0.95	0.93	0.56	0.90
4	1.07	0.88	0.83	1.33	1.91	0.75
5	1.18	0.97	0.80	1.24	-	-

(二) 参数估计与结果分析

泊松回归中所估计模型为 $\log(\lambda_{ij}) = \mu + \alpha_i + \beta_j + \gamma$ 和 $\log(\lambda_{ji}) = \mu + \alpha_j + \beta_i$, 以 2012 年数据为样本, 利用上述模型, 利用 2011—2012 赛季前 370 场比赛进球数据估计出 20 支队伍各自的参数, 似然函数为:

$$L(\alpha_i, \beta_j, \gamma, \mu, i=1, L, n) = \prod_{k=1}^N \frac{\exp(-\lambda_{ij}) \lambda_{ij}^{x_{ij}}}{x_{ij}!} \frac{\exp(-\lambda_{ji}) \lambda_{ji}^{y_{ji}}}{y_{ji}!}$$

表 5 2012 年各队参数估计结果

球队	平均进攻能力 α	标准差	平均防守能力 β	标准差
阿森纳	0.72	1.74	1.41	7.02
阿斯顿维拉	-0.69	1.83	1.52	4.56
布莱克本	0.07	0.96	1.98	4.72
博尔顿	0.09	0.72	1.97	4.74
切尔西	0.54	1.24	1.02	3.09
埃弗顿	-0.63	3.66	-2.39	8.10
富勒姆	-0.42	2.38	1.47	4.52
利物浦	-0.47	1.99	-6.45	29.35
曼城	0.96	1.67	-10.49	24.32
曼联	0.93	1.81	0.72	3.47
纽卡斯尔	0.36	1.23	1.46	4.73
诺维奇	0.09	1.25	1.82	4.85
QPR	-0.28	1.40	1.82	4.89
斯托克城	-2.04	5.61	1.48	4.53
桑德兰	-0.01	0.78	0.91	3.25
斯旺西	0.04	1.17	1.61	5.00
热刺	0.53	1.08	-5.24	19.89
西布罗姆维奇	0.24	1.74	1.60	5.07
维根	0.06	0.93	1.76	4.95
狼队	-0.07	0.99	2.01	4.76

注: 为使计算结果具有可比性, 估计过程中对参数施加约束条件 $\sum \alpha_i = 0, \sum \beta_j = 0$

以模型所估计 α, β, γ 和 μ 结果对各队的均值参数 λ 可计算并得出各队的实力水平, 主场效应参数值为 $\gamma = 1.759$, 该参数

说明比赛中主场因素对比赛结果有重要影响,Leighton 等人(2005)的研究也表明博彩商有动机利用这种主场效应和狂热球迷的“投资者情绪”来获利。利用前 370 场(即前 19 轮比赛)的数据得到的参数,表 6 给出最后 10 场比赛预测结果,发现该模型可有效预测最后的比分。

表 6 比赛结果预测

主场球队	客场球队	主队进球	客队进球	比分预测结果	胜负预测结果
切尔西	布莱克本	2	1	(2.57, 0.11)	对
埃弗顿	纽卡斯尔	3	1	(0.40, 0.00)	对
曼城	QPR	3	2	(3.35, 0.00)	对
诺维奇	阿斯顿维拉	2	0	(1.04, 0.11)	对
斯托克城	博尔顿	2	2	(0.19, 0.17)	对
桑德兰	曼联	0	1	(0.35, 0.23)	错
斯旺西	利物浦	1	0	(0.00, 0.11)	错
热刺	富勒姆	2	0	(1.28, 0.00)	对
西布罗姆维奇	阿森纳	2	3	(0.90, 0.36)	错
维根	狼队	3	2	(1.36, 0.19)	对

五、基本结论

- 第一,胜赔、平赔、负赔对比赛胜平负结果各自的预测是系统无偏差的有力证明了弱势市场有效性的存在。
- 第二,根据历史数据所建 logistic 模型,就 2012 年总体的预测效果来说,预测的准确度达到了 54.55%,预测效果较为理想。
- 第三,在高预测概率段,logistic 模型给出的交易策略的期望收益向 0 回归,该情况下考虑交易费用的存在模型可能给出正收益策略。
- 第四,泊松回归结果表明博彩商有动机利用这种主场效应和狂热球迷的“投资者情绪”来获利。模型准确地预测了比赛的进球数结果,为各队之间的实力对比提供了有力参考。

(上接第 101 页)态旅游、历史文化旅游和民俗文化旅游相结合。在加快南梁革命胜地建设的同时,要加快建设环县山城堡战役遗址纪念馆、河莲湾陕甘宁省委、省政府旧址扩建、曲子陇东分区习仲勋旧居、山岔红军长征宿营地、正宁王录邓小平旧居、习仲勋旧居恢复工作,逐步将资源优势转化为产业优势。

二要农旅结合。将红色旅游与观光农业旅游、文明村(镇)建设结合起来,在村庄整治中处处注重特色,注重加入更多的旅游元素。大力发展农家乐、橡皮坝垂钓、快艇、游泳休闲游和乡村游,带动黑木耳、黄花菜、白瓜子、土豆、玉米系列农产品开发,使之成为农村经济发展新的增长点。

三要赋予内涵。要善于挖掘、保护、利用地域特色文化,创新景区文化展示载体,充分体现革命老区的独特魅力。要融入具有庆阳特色的皮影、剪纸、香包、刺绣等非物质文化遗产,增强文化表现力和旅游内涵。

(三)区域联合,共同培育红色旅游品牌

加强区域合作,推进资源共享,已成为提升现代旅游业竞争力的必然趋势。《延安—庆阳红色旅游区域协作宣言》是庆阳市充分利用“陕甘宁”大区域优势,主动向全国红色旅游核心城市延安靠拢,积极开拓红色旅游市场,大力发展红色旅游产业的重要举措。延安与庆阳山水相连,两地民众自古以来就交往不断,具有厚重的历史人文情感,革命时期同属陕甘宁革命根据地的重要组成部分,有着共同的红色印记,两地人民相互支持、相互协作,共同促进了革命事业的成功。在共同发展红色旅游的今天,更应携起手来,加强协作,把两市的线路串连,让游

【参考文献】

[1] Mark J.Dixon ,Stuart G.Coles.Modeling Association Football Scores And Inefficiencies In The Football Betting Market[J].Applied Statist ,1997 ,46 (2) :265- 280.

[2] Graham ,H.Stott.Predicting bookmaker odds and efficiency for UK football [J].Applied Economics ,2008(40) :99- 109.

[3] Linda M.Woodland ,Bill M.Woodland.Market efficiency and the favorite longshot bias :the baseball betting market[J].Journal Of Finance ,1994 ,49 (1) :269- 279.

[4] Philip K.Gray ,Stephen F.Gray.Testing market efficiency :evidence from the NFL Sports Betting Market[J].Journal Of Finance ,1997 ,52(4) :1725 - 1737.

[5] MichaelCain ,David Law ,David Peel.The favorite- longshot bias and market efficiency in UK football betting[J].Scottish Journal of Political Economy ,2000 ,47(1) :25- 36

[6] Steven D.Levitt.Why are gambling markets organized so differently from Financial market?[J].Economic Journal ,2004 ,114 :495 -223- 246.

[7] DimitrisKarlis ,LoannisNtzoufras.Analysis of sports data by using bivariate Poisson models[J].Journal of the Royal Statistical Society ,2003 ,52(3) :381- 393.

[8] Daan Van Gemert.Modelling the Scores of Premier League Football Matches[J].Econometrics ,2010 ,18(67).

[9] Leighton Vaughan Williams.The Economics Of Gambling[M].USA and Canada :Routledge ,2002.

[10] Leighton Vaughan Williams.Information Efficiency in Financial and betting markets [M].United States of America :Cambridge University Press ,2005.

(责任编辑 X 校对 Z)

客全面地了解革命历史、党的创业史,更好地发挥三大教育基地的作用。突破行政区域界限,积极开展与周边延安、固原等红色旅游区的协作,推动设立陕甘宁无障碍红色旅游协作区,形成“资源互享、优势互补、游客互送”的红色旅游区域合作格局,共同合力打造“陕甘宁革命根据地红色旅游品牌”,切实增强庆阳市红色旅游的影响力。

(四)促销宣传,提高市场竞争力

- 一要强力促进市场营销。进一步加大对外宣传力度,在媒体的集中报道、重点市场的促销、节庆活动的举办、旅游网站建设等方面采取切实可行的措施,不断提高庆阳的知名度、美誉度和影响力。
- 二要深度开发旅游商品。重点开发具有庆阳特色的香包、皮影、剪纸、刺绣等旅游商品,搞好产品包装和商标注册,提升旅游商品的形象和质量,推动旅游产业化发展,提高旅游业的整体效益。

【参考文献】

[1] 许汉琴.革命老区红色旅游发展对策[J].当代经济 ,2011(6).

[2] 杨萍.区域旅游基础性研究[J].学术探索 ,2003(5)

[3] 王艳平 ,郭舒.旅游规划学[M].北京 :中国旅游出版社 ,2007.

[4] 沈中印.论大旅游视野下的产业互动与整合[J].商场现代化 ,2005 (30).

[5] 刘秉政.关于深度推进庆阳红色资源开发的思考[J].庆阳论坛 ,2011(2).

(责任编辑 X 校对 Z)